19

Analyzing More Difficult Tables

		Exh.	Page
Introduction 19-1			
19A.	Another Nebraska example -- Nebraska voting 2 19-1	A	
		1	19-3
		2	19-4
19B.	Polishing the new example 19-8	3	19-5/7
	- checking ways 19-16	B	
		C	
		4	19-9
19C.	Going on with the example 19-19	5	19-10/1[illegible]
	- repair and go on 19-20	6	19-13/1[illegible]
	- looking again 19-20	7	19-17/[illegible]
		8	19-21/[illegible]
19D.	Relax, relax 19-26	9	19-24
	- stopping sooner, strip by stripe 19-26	D	
	median, at 19-31	10	19-27/[illegible]
	- judgment often softens further 19-32	E	
	- checking should be relaxed 19-32	11	19-34/[illegible]
		12	19-37
19E.	Tables with holes 19-33	13	19-38/[illegible]
	- after an analysis 19-33	F	
	- from the beginning 19-3[illegible]	14	19-41/[illegible]
		15	19-43/[illegible]
19F.	More holes may make more trouble 19-40	G	
	- a clearer example 19-47	16	19-45
		17	19-49
19G.	Differences can help 19-47	18	19-51/5[illegible]
	- adjacent columns 19-47	19	19-53
	- other choices 19-50	H	
		K	
	**********	20	19-59
		P	
19H.	How far have we come? 19-54		
	**********		
19K.	Uniqueness? (Optional background) 19-57		
	- first, means 19-57		
	- now, medians 19-58		
	- the morals 19-58		
	**********		

19

Analyzing more difficult tables

Already in chapter 16 we developed techniques for resistant two-way analysis. In chapter 17, where the emphasis is different, we said nothing about resistance.

It is time that we looked at an actual moderately large table where resistance is clearly needed. After that it will be time to learn how to keep the labor of arithmetic as small as we know how to do.

On the way to holding down arithmetic we will want to look at multiple possibilities for two-way analyses of the same data, all with all stripe medians zero. (We will decide not to let this bother us.)

19A. Another Nebraska example -- Nebraska voting 2

The famous election district shaped like a salamander and attributed (unfairly, it is now believed) to Elbridge Gerry produced the word "gerrymander". This word was used to emphasize the care with which that district was shaped to a political end. Chapter 17's group of ten contiguous Nebraska counties, as our maps -- particularly exhibit 30 of chapter 17 -- show, is shaped more like a poodle than like a salamander, but there is rightful ground for suspicion that they were picked for a particular data-analytical purpose.

Those who have not yet learned from experience how often one can peel off layer after layer of appearances and cues may think that our 10 counties were selected so that there would be layer after layer. This is not so.

Those 10 counties were selected to be unusually well behaved and consistent, so much so as to allow effective analysis by means. When we take a different set -- though we may well need analysis by some more resistant method -- we will still be likely to find layer after layer of appearances. Our selection was to keep the analysis simple, not to change its layered character.

Another example, also from Nebraska's presidential voting, may help to convince all of us that it is easy to find data that needs a more resistant elementary analysis than the use of row and column means will provide. Exhibit 1 maps the counties. Exhibit 2 sets out the raw data.

The first panels of exhibit 3 set out the elementary analysis by means. Panel C shows the residuals separated according to whether or not they are outside. As the lower half of this panel shows, while residuals of 9 or 10 are already outside, there are residuals of 14 and even 25. Does the presence of such large residuals indicate that the use of means as a basis for our analysis is unwise?

Eleven more counties in Nebraska

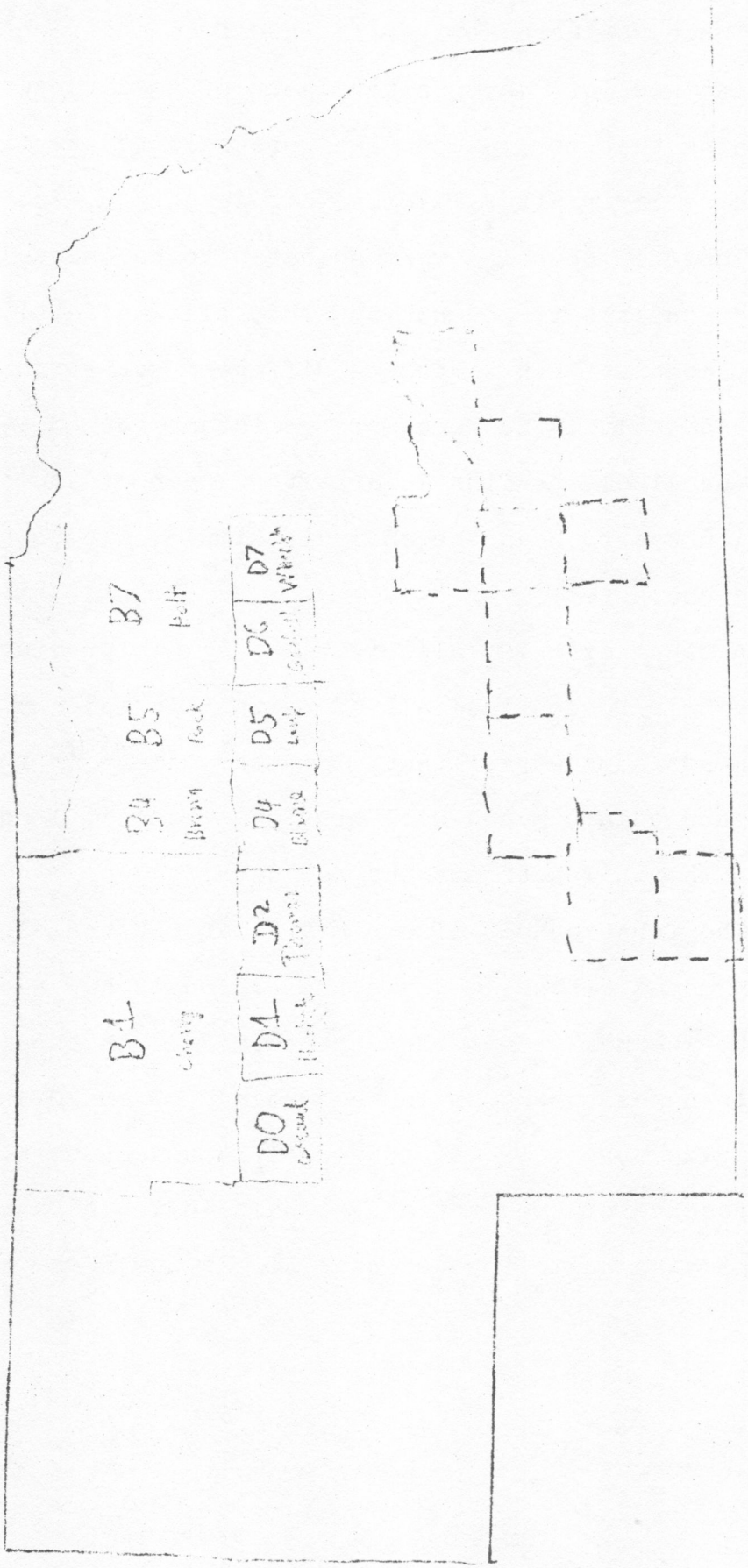

exhibit 2 of chapter 19 -- Nebraska voting 2

Raw % Democratic for 11 counties in 12 Presidential elections

(see exhibit 1 for map locations)

A) DATA

County	'20		'28		'36		'44		'52		'60	
D0	35.3	35.8	58.9	75.7	54.4	36.5	34.5	33.7	18.9	16.7	23.6	39.6
D1	32.3	25.2	23.6	66.9	39.6	26.7	23.8	25.7	14.9	14.8	13.8	29.0
B1	28.8	30.2	30.5	61.9	51.0	39.7	37.2	41.1	23.4	26.8	27.9	38.9
D2	37.9	37.2	27.0	60.6	49.7	36.3	38.8	43.3	19.6	22.3	25.1	39.9
D4	34.2	22.6	26.4	62.6	51.0	40.7	40.4	49.6	23.0	26.4	22.2	37.4
B4	27.0	29.1	24.7	56.9	45.0	35.4	32.5	37.4	21.8	25.9	22.9	41.0
D5	22.8	17.7	15.0	55.3	42.6	34.9	27.2	47.2	17.7	24.0	22.5	33.6
B5	27.0	23.7	22.7	56.1	42.5	35.2	34.0	36.0	17.9	23.2	18.9	31.0
D6	26.5	19.6	16.5	54.7	47.2	33.6	31.3	43.6	19.5	21.9	22.6	38.8
B7	32.2	25.7	45.4	66.1	51.3	38.4	37.9	45.4	25.3	30.7	37.0	46.2
D7	27.0	19.1	35.2	77.6	52.6	46.3	44.2	55.3	33.7	35.8	36.0	43.9

B) SOURCE

Richard M. Scammon 1965. America at the Polls. A handbook of American presidential election statistics 1920-1964. Governmental Affairs Institute. University of Pittsburgh Press. From pages 276 to 287.

exhibit 3 of chapter 19 -- Nebraska voting 2

Elementary analysis by means for the data of exhibit 2

A) The ANALYSIS

County	'20		'28		'36		'44		'52		'60		eff	fit
DO	1.5	6.1	25.6	8.9	2.9	-3.9	-3.9	-11.7	-6.3	-11.4	-4.9	-2.2	3.7	38.6
D1	9.1	6.1	1.1	10.7	-1.3	-3.1	-4.0	-9.1	.3	-2.7	-4.1	-2.2	-6.9	28.0
B1	-2.9	2.6	-0.7	-2.8	1.6	1.4	.9	-2.2	.3	.8	1.5	- .8	1.6	36.5
D2	6.2	9.6	-4.7	-4.1	.3	-2.0	2.5	.0	-3.5	-3.7	-1.3	.2	1.6	36.5
D4	2.6	-4.9	-4.7	-2.0	1.7	2.5	4.2	6.4	.0	.5	-4.1	-2.2	1.5	36.4
B4	-1.3	4.7	-3.3	-4.6	-1.2	.3	- .6	-2.7	1.9	3.1	- .3	4.5	-1.6	33.3
D5	-2.4	-3.4	-9.7	-2.9	- .3	3.1	-2.6	10.4	1.1	4.5	2.6	.4	-4.9	30.0
B5	1.1	1.5	-2.7	-2.8	-1.1	2.7	3.5	-1.5	.6	3.0	-1.7	-2.9	-4.2	30.7
D6	.0	-2.8	-9.5	-4.8	3.0	.5	.2	5.5	1.6	1.1	1.4	4.3	-3.6	31.3
B7	-3.0	-5.4	10.7	-2.1	-1.6	-3.4	-1.9	-1.4	-1.3	1.2	7.1	3.0	-5.5	29.4
D7	-10.4	-14.2	-1.9	7.2	-2.5	2.3	2.2	6.3	4.5	4.1	3.9	-1.5	7.3	42.2
eff	-4.8	-8.9	-5.3	28.2	12.9	1.8	-0.2	6.8	-13.4	-10.5	-10.1	3.2	34.9	00.0
fit	30.1	26.0	29.6	63.1	47.8	36.7	34.7	41.7	21.5	24.4	24.8	38.1	00.0	-34.9

(Cont'd)

exhibit 3 (cont'd.)

B) The 132 RESIDUALS -- unit 0.1%D

High*	Low*	To		
256	-104	4	L	LLLL
107	-142	7	-9	751
107	-117		-8	
104	-114		-7	
(96)	(-97)	8	-6	3
(91)	(-95)	9	-5	4
(89)	(-91)	19	-4	9771580911
		28	-3	043914957
		47	-2	94878098150527722 29
		60	-1	3932169543375
		(7)	-0	7360038
		65	0	0335923368524
		52	1	5151674916125 4
		33	2	669573526
		29	3	0151090
		22	4	7255153
		15	5	5
		14	6	21143
		9	7	21
		7	8	9
		6	9	16
		4	H	HHHH

60 + 7 + 65 = 132, hinges at 33h

Hinges:	-27h	and +26
Midspread:		53h
Sides:	-81	and +79
Outside:	7	and 7
(count/20):		6h

*Entries in parentheses also appear in main display, shown here as outside.

C) THE RESIDUALS SEPARATED -- outside, and NOT outside

County	Election '20		'28		'36		'44		'52		'60	
residuals NOT outside												
DO	1.5	6.1	.	.	2.9	-3.9	-3.9	.	-6.3	.	-4.9	-2.2
D1	.	6.1	1.1	.	-1.3	-3.1	-4.0	.	.3	-2.7	-4.1	-2.2
B1	-2.9	2.6	-0.7	-2.8	1.6	1.4	.9	-2.2	.3	.8	1.5	-0.8
D2	6.2	.	-4.7	-4.1	.3	-2.0	2.5	.0	-3.5	-3.7	-1.3	.2
D4	2.6	-4.9	-4.7	-2.0	1.7	2.5	4.2	6.4	.0	.5	-4.1	-2.2
B4	-1.3	4.7	-3.3	-4.6	-1.2	0.3	-0.6	-2.7	1.9	3.1	-0.3	4.5
D5	-2.4	-3.4	.	-2.9	-0.3	3.1	-2.6	.	1.1	4.5	2.6	.4
B5	1.1	1.5	-2.7	-2.8	-1.1	2.7	3.5	-1.5	.6	3.0	-1.7	-2.9
D6	.0	-2.8	.	-4.8	3.0	0.5	.2	5.5	1.6	1.1	1.4	4.3
B7	-3.0	-5.4	.	-2.1	-1.6	-3.4	-1.9	-1.4	-1.3	1.2	7.1	3.0
D7	.	.	-1.9	7.2	-2.5	2.3	2.2	6.3	4.5	4.1	3.9	-1.5
residuals outside, rounded to nearest 1%												
DO	.	.	26	9	.	.	.	-12	.	-11	.	.
D1	9	.	.	11	.	.	.	-9	.	.	.	.
B1	.	.	.	.	.	.	.	.	.	.	.	.
D2	.	10	.	.	.	.	.	.	.	.	.	.
D4	.	.	.	.	.	.	.	.	.	.	.	.
B4	.	.	.	.	.	.	.	.	.	.	.	.
D5	.	.	-10	.	.	.	.	10	.	.	.	.
B5	.	.	.	.	.	.	.	.	.	.	.	.
D6	.	.	-10	.	.	.	.	.	.	.	.	.
B7	.	.	11	.	.	.	.	.	.	.	.	.
D7	-10	-14	.	.	.	.	.	.	.	.	.	.

exhibit 3 of chapter 19 (cont'd.)

D) ALL the RESIDUALS SHOWN GRAPHICALLY -- and the effects in open stem-and-leaf to the nearest whole %.

	'20		'28		'36		'44		'52		'60	
D0	.	x	X	X	x	o	o	0	o	0	o	.
D1	X	x	.	X	.	o	o	0	.	.	o	.
B1	o	.	.	o	.	.	.	.	.	.	.	.
D2	x	X	o	o	.	.	.	.	o	o	.	.
D4	.	o	o	.	.	.	x	x	.	.	o	.
B4	.	x	o	o	.	.	.	.	.	x	.	x
D5	.	o	0	o	.	x	.	X	.	x	.	.
B5	.	.	.	o	.	.	x	.	.	x	.	o
D6	.	.	0	o	x	.	.	x	.	.	.	x
B7	o	o	X	.	.	o	.	.	.	.	x	x
D7	0	0	.	X	.	.	.	x	x	x	x	.

	-0	0
		4
	7	
		2
		2
		2
	2	
	5	
	4	
	4	
	6	
		7

2				8								
1					2							
0						2		7				3
-0	5	9	5				0					
-1									3	0	0	

19-7a

19B. Polishing the new example

We ought to think harder about the one residual of nearly 26. How much distortion of means does such a value reflect? An easy way to learn is to look at a similar analysis of another table of the same size, a table in which all entries but one are alike. For convenience we choose zero for the value of the 131 equal entries and a value large enough to make the residual close to 25.6 for the other entry. Exhibit 4 shows the result. Even in a table as large as 11-by-12, one such stray value pushes 23 other residuals off more than 2%. If we plan to go further with our analyses, it is clearly important to replace our analysis by means by a much more resistant two-way analysis.

One way to do this is to median polish the analysis we already have. Exhibits 5 and 6 show the first (whole of medians) and second (halves of medians) full steps involved in doing this in a careful and balanced way.

We have tried to simplify the calculation of the medians as much as we can. Look at the '20 column of panel A of exhibit 5, where the first steps were to make three counts among the 11 residuals in this stripe. First for those that are -, of which 5 were found. Then, for those that are 0. Last, for those that are +. Clearly, for this column the median is zero. For the next column, with counts 5, 0, 6, the median is the smallest + residual, which turns out to be +15. For the third column -- with counts 8, 0, 3 -- the

exhibit 4 of chapter 19 -- Nebraska voting 2

Two-way analysis by means of a dead level table with one stray

A) The DATA -- made up for the purpose

0	0	31.0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

B.) The ELEMENTARY ANALYSIS -- by MEANS

-2.4	-2.4	25.6	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	2.4	2.6
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
0.2	0.2	-2.6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	-0.2	0.0
-0.2	-0.2	2.6	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.2	0.0
0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.2

Exhibit 5 of chapter 19 -- Nebraska voting 2

Median polishing the residuals of exhibit 3 -- first step and results

A) MEDIANS for COLUMNS -- 5, 1, 5 ideal

Election	'20		'28		'36		'44		'52		'60	
Count -	5	5	8	8	6	4	5	6	3	3	6	6
Count 0	1	0	0	0	0	0	0	1	1	0	0	0
Count +	5	6	3	3	5	7	6	4	7	8	5	5
Median	.0	1.5	-2.8	-2.8	-.3	.5	.2	-1.4	.3	1.1	-.3	-.8
(check)			✓	✓		✓	✓	✓	✓	✓	✓	✓

For examples, see text.

B) MEDIANS for ROWS -- 5, 1, 1, 5 ideal

County	D0	D1	B1	D2	D4	B4	D5	B5	D6	B7	D7
Count +	6	7	5	6	5	6	5	5	3	8	5
Count 0	0	0	0	1	1	0	0	0	1	0	0
Count -	6	5	7	5	6	5	6	6	8	3	7
(m)	(1.5)	(-1.5)	(.3)	(-1.3✓)	(.0)	(-.3)	(-.3)	(-1.1)	(-.5)	(-1.6✓)	(2.2)
(m)	(-2.2)	(-2.2✓)	(.8✓)	(.0)	(.5✓)	(-.6✓)	(.4)	(.6)	(1.1✓)	(-1.4)	(2.3✓)
Median	-.4	-1.8	.6	-.6	.2	-.4	0	-.2	.8	-1.5	2.2

C) FIRST ADJUSTMENTS -- body values in 0.1%D

	.0	1.5	-2.8	-2.8	-.3	.5	.2	-1.4	.3	1.1	-.3	-.8
-.4	-4	11	-32	-32	-7	1	-2	-18	-1	7	-7	-12
-1.8	-18	-3	-46	-46	-21	-13	-16	-32	-15	-7	-21	-26
.6	6	21	-22	-22	3	11	8	-8	9	17	3	-2
-.6	-6	9	-34	-34	-9	-1	-4	-20	-3	5	-9	-14
.2	2	17	-26	-26	-1	7	4	-12	5	13	-1	-6
-.4	-4	11	-32	-32	-7	1	-2	-18	-1	7	-7	-12
0	0	15	-28	-28	-3	5	2	-14	3	11	-3	-8
-.2	-2	13	-20	-20	-5	3	0	-16	1	9	-5	-10
.8	-8	23	-20	-20	5	13	10	-6	11	19	5	0
-1.5	-15	0	-43	-43	-18	-10	-13	-29	-12	-4	-18	-23
2.2	22	37	-6	-6	19	27	24	-8**	25	33	19	14

Examples: -4 from -0.4 = .0 + (-.4), -18 from -1.8 = 0 + (-1.8).

** to be explained in exhibit 7.

exhibit 5 (cont'd.)

D) FIRST-STEP POLISHED -- effect analysis in 0.1%D

County	'20		'28		'36		'44		'52		'60		eff
D0	19	50	288	117*	36	-40	-37	-99	-62	-121	-42	-10	33
D1	109	64	57	153	8	-18	-24	-59	18	-20	-20	4	-87
B1	-35	5	15	-6	13	3	1	-14	12**	-9	12	6**	22
D2	68	87	-13	-7	12	-19	29	20	-38	-42	-4	16	10
D4	24	-66	-21	6	18	18	38	76	-5	-8	-40	-16	13
B4	-9	36	-1	-14	-5	2	-4	-9	20	38**	4	57	-20
D5	-24	-49	-69	-1	0	26	-28	118	8	56**	29	12	-49
B5	13	2	3	2	-6	24	35	1	5	21	-12	-19	-44
D6	-8	-51	-75	-28	25	-8	-8	61	5	-8	9	43	-28
B7	-15	-54	150	22	2	-24	-6	15	-1	16	89	53	-70
D7	-126	-179	-13	78	-44	-4	-2	71	20	8	29	-29	95
	-48	-74	-81	310**	126	23	0	54	-131	-94	-98**	24	349

Examples: 19 = 15 - (-4) with 15 from 1.5, 109 = 91 - (-18), with 91 from 9.1.
33 = 37 + (-4) with 37 from 3.7 and -4 from 0.4,
-87 = -69 + (-18) with -69 from -6.9 and -18 from -1.8.
-48 = -48 + 0 with -48 from -4.8 and 0 from .0,
-74 = -89 + 15 with -89 from -8.9 and 15 from 1.5.

* To be explained in exhibit 7.
**To be explained in exhibit 7.

(cont'd.)

exhibit 5 (cont'd.)

E) STEM-and-LEAF for RESIDUALS -- of panel D

|High*|	|Low*|	|To|		
		8	H	see column, count = 8
109	-126	11	7	861
87	-179	14	6	841
288	-99	19	5	07673
150	-121	20	4	3
117	(-75)	24	3	6688
153	(-69)	36	2	425649000190
118	(-66)	50	1	93532885826262
89	(-62)	(20)	0	52362802321185584946
(78)	(-59)	62	-0	98167156844862951088 4
(76)		41	-1	53348942069
(71)		30	-2	418448009
(68)		21	-3	5758
(64)		17	-4	940220
(61)		11	-5	149
		8	-6	692
		5	-7	5
		4	L	see column, count = 4

50 + 20 + 62 = 132, hinges at 33h

Hinges:	+ 20 and - 17
Midspread:	37
Sides:	+ 57 and - 54
Outside:	14 and 9
(count/20):	6h

* Entries in parenthesis are in main stem-and-leaf, but are outside.

Compare: Midspread in exhibit 3: 53h
Outside in exhibit 3: 7 and 7

exhibit 6 of chapter 19 -- Nebraska voting 2

Median polishing the residuals of exhibit 3
-- second full step (halves of medians) and results
(all in 0.1%D)

A) MEDIANS for COLUMNS -- 5, 1, 5 ideal

Election	'20		'28		'36		'44		'52		'60	
Count -	6	5	6	5	3	6	7	4	4	6	5	4
Count 0	0	0	0	0	1	0	0	0	0	0	0	0
Count +	5	6	5	6	7	5	4	7	7	5	6	7
Median	-8	2	-1	2	8	-4	-4	15	5	-8	4	6
					✓		✓	✓	✓			✓

B) MEDIANS for ROWS -- 5, 1, 1, 5 ideal

County	D0	D1	B1	D2	D4	B4	D5	B5	D6	B7	D7
Count -	7	5	4	6	6	6	5	3	7	5	7
Count 0	0	0	0	0	0	0	1	0	0	0	0
Count +	5	7	8	6	6	6	6	9	5	7	5
(m	(-37✓	(4	(3	(-4	(-5	(-1	(0	(2	(-8	(2	(-4
(m	(-10	(8✓	(5✓	(12	(6	(2	(8	(3✓	(-8	(16	(-2
Median	-24	6	4	4	0	0	4	2	-8	9	-3

C) SECOND ADJUSTMENTS -- using halves of medians

	-4	1	0	1	4	-2	-2	8	2	-4	2	3
-12	-16	-11	-12	-11	-8	-14	-14	-4	-10	-16	-10	-9
3	-1	4	3	4	7	1	1	11	5	-1	5	6
2	-2	3	2	3	6	0	0	10	4	-2	4	5
2	-2	3	2	3	6	0	0	10	4	-2	4	5
0	-4	1	0	1	4	-2	-2	8	2	-4	2	3
0	-4	1	0	1	4	-2	-2	8	2	-4	2	3
2	-2	3	2	3	6	0	0	10	4	-2	4	5
1	-3	2	1	2	5	-1	-1	9	3	-3	3	4
-4	-8	-3	-4	-3	0	-6	-6	4	-2	-8	-2	-1
4	0	5	4	5	8	2	2	12	6	0	6	7
-2	-6	-1	-2	-1	2	-4	-4	6	0	-6	0	1

(cont'd.)

19-13

exhibit 5 (cont'd)

D) MEDIAN POLISHED -- effect analysis in 0.15%

County	'20		'28		'36		'44		'52		'60		eff	eff*
D0	35	61	300	128	44	-26	-23	-95	-52	-105	-32	-1	21	1
D1	110	60	54	149	1	-19	-25	-70	13	-19	-25	-2	-84	-69
B1	-33	2	13	-9	7	3	1	-24	8	-7	8	1	24	44
D2	70	84	-15	-10	6	-19	29	10	-42	-40	-8	11	12	32
D4	28	-67	-21	5	14	20	40	68	-7	-4	-42	-19	13	33
B4	-5	35	-1	-15	-9	4	-2	-17	18	42	2	54	-20	0
D5	-22	-52	-71	-4	-6	26	-28	108	4	58	25	7	-47	-27
B5	16	0	2	0	-11	25	36	-8	2	24	-15	-23	-43	-23
D6	0	-48	-71	-25	25	-2	-2	57	7	0	-11	44	-32	-12
B7	-15	-59	146	17	6**	-26	-8	3	-7	16	83	46	-66	-46
D7	-120	-178	-11	79	-46	0	2	65	20	14	26	-30	93	73
eff	-52	-73	-81	311	130	21	-2	62	-129	-98	-96	27	349	329
eff*	-25	-46	-54	338	157	48	25	89	-102	-71	-69	54	322	302

eff* is the result of median polishing the effects.

Examples:

** to be explained in exhibit 7.

(cont'd)

exhibit 6 (cont'd.)

E) STEM-and-LEAF for RESIDUALS -- of panel D

High*	Low*	|To|		
|High*|	|Low*|	8	H	see column, count = 8
110	-120	10	7	09
84	-178	14	6	1085
300	-95	18	5	4784
146	-105	23	4	40246
128	(-71)	26	3	556
149	(-71)	36	2	8506590450
108	(-70)	46	1	6374038641
83	(-67)	(23)	0	022501766341238427082 17
(79)	(-59)	63	-0	501949620228877748 12
(70)		43	-1	5510519979519
(68)		30	-2	21556358453
(65)		19	-3	320
		16	-4	86202
		11	-5	292
		8	-6	7
		7	-7	110
		4	L	see column, count = 4

46 + 23 + 63 = 132, check

hinges at depth 33h

Hinges:	+22 and -19
Midspread:	41
Sides:	+63 and -60
Outside:	12 and 8
(comparison par):	6h

*Entries in parenthesis also in main stem-and-leaf, but are outside.

COMPARISON

	Midspread	Outside
Exhibit 3	53h	7 and 7
Exhibit 5	37	14 and 9
Above	41	12 and 9

19-15

median is negative, but is not the smallest negative residual. Working with the eight negative residuals, we find the residual of depth 6 to be -2.8. Then we check by counting the number of residuals more positive than -2.8, finding 5 as we should. Hence the check mark. And so on. As the foot of panel E of exhibit 6 shows, the result is a very considerable improvement in the analysis, with the midspread shrinking to no more than 3/4 of its former value.

How much has the analysis itself changed? Exhibit 7 sets out to answer this question by writing down the difference analysis. Once we do this, it is important to check the arithmetic.

* checking pays *

Exhibit 7 gives the raw material for checking the possibility of our arithmetic. We could do this exactly as we explained in chapter 16. For variety, let us take a slightly different route, finding differences within rows, but -- except for extra differences to settle error locations -- only for successive pairs of columns.

Looking down the first two columns of panel A, we find all ('20) - ('24) differences the same, mainly +20. Taking the next two columns, we find all ('28) - ('32) differences

exhibit 7 of chapter 19 -- Nebraska voting 2

Change in analysis from exhibit 3 to exhibit 6
(in 0.1%D)

A) The DIFFERENCE ANALYSIS

County	'20		'28		'36		'44		'52		'60		eff	check sum
D0	20	0	44	39*d	15	13	16	22	11	9	17	21	36**	-15
D1	19	-1	43	42	14	12	15	21	10	8	16	20	5	20
B1	-4	-24	20	19	-9	-11	-8	-2	5**	-15	-7	9**	28	37
D2	8	-12	32	31	3	1d	4	10	-7	-3	5	9	16	25
D4	2	-18	26d	25	-3	-4	-2	4	-7	-0	-1	3	18**	21
B4	8	-12	32	31	3	1	4	10	-1d	11**	5	9	16	25
D5	2	-18	26	25	-3	-5	-2	4	-7	13**	-1	3	22	25
B5	5	-15	29	28	0	-2	1	7	-4	-6	2	6	19	25
D6	0	-20	24	23	-5	-7	-4	2	-0	-11	-3	1	24	25
B7	15	-5	39	38d	22**	8	11	17	6	4	12	16	9	25
D7	-16	-36	8	7	-21	-23	-20	2**	-25	-27	-19	-15	0**	-15
eff	25	43	-1	56**	28	30	27	21	32	34	32**	22	-47	-25
check sum	7	7	7	63	7	7	7	23	7	7	13	37	-47	

* First failure to check. Difference ('28) - ('32) is +1 elsewhere in column. ('28 D4 had error in finding difference.)

** Other failures to check.

d indicates difference calculated wrong at first attempt.

Examples: 20 = 35 - 15, 15 from 1.5, 19 = 110 - 91, 91 from 9.1.

B) PROBLEMS

1) Make necessary corrections to exhibits 5, 6 and 7, eliminating failures to check.

2) Using corrected values for exhibit 6, check its relation to exhibit 2 by making a difference analysis.

(cont'd.)

19-17

exhibit 7 (cont'd.)

C) STEM-and-LEAF for the RESIDUALS of panel A

```
 |tc|
   3        4|432
  10        3|2299118
  25        2|006,694,558,322,110
  45        1|959,543,265,107,101,376,26
 (36)       0|828,250,873,311,844,144,722,569,345,529,939,361
  51       -0|401,593,305,552,782,242,771,749,396,711,3
  20       -1|628,285,105,195
   8       -2|4013057
   1       -3|6
           -4|
```

45 + 36 + 51 = 133, hinges at 33h

Hinges:	+15	and - 4h
Midspread:		19h
Sides:	+34	and -24
Outside:	6	and 3
(count/20):		6h

+1 except the first. Which is in error? Looking at the first few ('24) - ('28) differences, we find them all -44, so the '28 entry for DO seems okay. Thus the '32 entry must be in trouble. And so on. A total of 5 residuals are in error.

The check sum columns find the other five errors. More frequent errors among effects -- five out of 24 as compared to 5 out of 132 among residuals -- is usual. This is probably not alone because the effects are often bigger numbers.

When making two-way analyses of any size by hand, we ought always to write out a difference analysis as a check.

19C. Going on with the example

As panel C of exhibit 7 shows, the midspread of the changes in residuals, namely 19h, is about half the size of the midspread of the polished residuals. This, too, makes the effect of polishing seem large.

Looking back at panel A of exhibit 7, we see that the differences due to polishing are not only substantial but fall in a rather systematic pattern. Further analysis is likely to be disturbed if it is based on the two-way analysis by means. We need a resistant analysis. First, however, we need a possible analysis.

* repair and go on *

We have emphasized in chapter 16 that checking, as already done in exhibit 7, should lead to repair, not to autoredimination. Exhibit 8 shows what happens when we repair exhibit 7 by first finding a corrected form of exhibit 7 and then, from this, a correct analysis based on exhibit 6. Panel D indicates the 8 changes in residuals.

Panel E repairs the stem-and-leaf display of panel E, exhibit 6. The stems are put off to the left, so that we can work on both sides of two vertical lines. Around the first vertical we place the old (to the left) and new (to the right) values of the changed residuals. Then we can copy the old stem-and-leaf display, dropping out characters for old values and adding them for new. As a final result, this example shows one less outside value. It is often convenient to repair a stem-and-leaf, rather than to repeat its construction.

* looking again *

What can we see in our example, once it has been resistantly analyzed? Exhibit 9 provides the possible, resistant version of panel C of exhibit 3. Exhibit 9a summarizes the situation more graphically. We can now see such things as:

- greater activity in the early elections and in 1948.

exhibit 8 of chapter 19 -- Nebraska voting 2

A correct analysis based on exhibit 6 by means of exhibit 7 (all in 0.1%D)

A) Panel A, exhibit 7 as CORRECTED -- altered values bold

County	'20		'28		'36		'44		'52		'60		eff	fit
D0	20	0	44	43	15	13	16	22	11	9	17	21	4	-16
D1	19	-1	43	42	14	12	15	21	10	8	16	20	5	-15
B1	-4	-24	20	19	-9	11	-8	-2	**-13**	-15	-7	-3	28	8
D2	8	-12	32	31	3	1	14	10	**-1**	-3	5	9	16	-4
D4	2	-18	26	25	-3	-5	-2	4	-7	-9	-1	3	22	2
B4	8	-12	32	31	3	1	4	10	-1	-3	5	9	16	-4
D5	2	-18	26	25	-3	-5	-2	4	-7	-9	-1	3	22	2
B5	5	-15	29	28	0	-2	1	7	-4	-6	2	6	19	-1
D6	0	-20	24	23	-5	-7	-4	2	-9	-11	3	1	24	4
B7	15	-5	39	38	**10**	8	11	17	6	4	12	16	9	-11
D7	-16	-36	8	7	-21	-23	-20	**-14**	-25	-27	-19	-15	40	20
eff	23	43	-1	0	28	30	27	21	32	34	**26**	22	-47	-27
fit	-4	16	-28	-27	1	3	0	-6	5	7	**-1**	-5	-20	0

B) PROBLEMS

1) State two simple comparisons that show that one of the altered values has been correctly altered.

2) Repeat this for the others.

Examples: (The fits are matched to the correct entries in panel D, exhibit 6.)

-4 = -52 - (-48), hence -27 = -4 - 23.
16 = -73 - (-89), hence -27 = 16 - 43, check.

(cont'd.)

exhibit 8 (cont'd.)

C) The ANALYSIS of panel D, exhibit 6, ALTERED to fit panel A above -- altered values **bold**

County	'20		'28		'36		'44		'52		'60		eff	eff*
D0	35	61	300	**132**	44	-26	-23	-95	-52	-105	-32	-1	**21**	**41**
D1	110	60	54	149	1	-19	-25	-70	13	-19	-25	-2	-84	-64
B1	-33	2	13	-9	7	3	1	-24	**-10**	-7	8	**-11**	**24**	**44**
D2	70	84	-15	-10	6	-19	29	10	**-36**	-40	-8	11	12	32
D4	28	-67	-21	-5	14	20	40	68	-7	-4	-42	-19	**37**	**57**
B4	-5	35	-1	-15	-9	4	-2	-17	18	**28**	2	54	-20	0
D5	--22	-52	-71	-4	-6	26	-28	108	4	**36**	25	7	-47	-27
B5	16	0	2	0	-11	25	36	-8	2	24	-15	-23	-43	-23
D6	0	-48	-71	-25	25	-2	-2	57	7	0	-11	44	-32	-12
B7	-15	-59	146	17	**-6**	-26	-8	3	-7	16	83	46	-66	-46
D7	-120	-178	-11	79	-46	0	2	**47**	20	14	20	-30	**93**	**113**
eff	-52	-73	-81	**255**	130	21	-2	62	-129	-98	**-102**	27	349	329
eff*	-25	-46	-54	**282**	157	48	25	89	-102	-71	**-75**	54	322	

Construction steps: 1) Draw large open circles -- indicating entries to be bold -- at the location of each value to be altered.

2) Fill in all other values unchanged.

3) Calculate altered values.

Examples: 35 = 15 + 20 (15 from exhibit 3, this as a check).
132 = 89 + 43.
-6 = -16 + 10.

Note that two row effects are unchanged, the error having been between eff and eff* .

(cont'd.)

exhibit 8 (cont'd.)

D) CHANGES in RESIDUALS

128 to 132
6 to -6
65 to 47
8 to -10
-42 to -36
42 to 28
58 to 36
1 to -11

E) REPAIR of STEM-and-LEAF

stem	old	new	to	results added bold
H	H	H	8	see column, count = 8 (*)
7			10	09
6	5		13	108
5	8		16	474
4	2	7	21	40467
3		6	25	5566
2		8	36	85065904508
1			46	6374038641
0	861		(20)	02250763412342708217
-0		6	66	50194962022387774812 6
-1		01	45	551051997951901
-2			30	21566358453
-3		6	19	3206
-4	2		15	8620
-5			11	292
-6			8	7
-7			7	110
L			4	see column, count = 4

46 + 20 + 66 = 132, hinges at 33h

Hinges:	+22 and -19
Midspread:	41
Sides:	+63 and -60
Outside:	11 and 8
(count/20):	(6h)

Note: There is also one change of value, from 128 to 132, in the "High" column.

exhibit 9 of chapter 19 -- Nebraska voting 2

The residuals of exhibit 8 separated

County	Election '20		'28		'36		'44		'52		'60		
	residuals NOT outside												
D0	3.5	6.1	.	.	4.4	-2.6	-2.3	.	-5.2	.	-3.2	- .1	
D1	.	6.0	5.4	.	.1	-1.9	-2.5	.	1.3	-1.9	-2.5	- .2	
B1	-3.3	.2	1.3	-.9	.7	.3	.1	-2.4	-1.0	-.7	.8	-1.1	
D2	.	.	-1.5	-1.0	.6	-1.9	2.9	1.0	-3.6	-4.0	-.8	1.1	
D4	2.8	.	-2.1	.5	1.4	2.0	4.0	.	-.7	-.4	-4.2	-1.9	
B4	-.5	3.5	-.1	-1.5	-.9	.4	-.2	-1.7	1.8	2.8	.2	5.4	
D5	-2.2	-5.2	.	-.4	-.6	2.6	-2.8	.	.4	3.8	2.5	.7	
B5	1.6	.0	.2	.0	-1.1	2.5	3.6	-.8	.2	2.4	-1.5	-2.3	
D6	.0	-4.8	.	-2.5	2.5	-.2	-.2	5.7	.7	.0	-1.1	4.4	
B7	-1.5	-5.9	.	1.7	-.6	-2.6	-.8	.3	-.7	1.6	.	4.6	
D7	.	.	-1.1	.	-4.6	.0	.2	4.7	2.0	1.4	2.0	-3.0	
	residuals outside, rounded to nearest 1%												(Count)
D0	.	.	30	13	.	.	.	-10	.	-10	.	.	(4)
D1	11	.	.	15	.	.	.	-7	.	.	.	.	(3)
B1	.	.	.	.	.	.	.	.	.	.	.	.	.
D2	7	8	.	.	.	.	.	.	.	.	.	.	(2)
D4	.	-7	.	.	.	.	.	7	.	.	.	.	(2)
B4	.	.	.	.	.	.	.	.	.	.	.	.	.
D5	.	.	-7	.	.	.	.	11	.	.	.	.	(2)
B5	.	.	.	.	.	.	.	.	.	.	.	.	.
D6	.	.	-7	.	.	.	.	.	.	.	.	.	(1)
B7	.	.	15	.	.	.	.	.	.	.	8	.	(3)
D7	-12	-18	.	8	.	.	.	.	.	.	.	.	(4)
(count)	(3)	(4)	(4)	(3)				(5)		(1)	(1)		((21))

exhibit 9a of chapter 19

The residuals simply coded -- in plotting style

	'20		'28		'36		'44		'52		'60	
DO	x	x	X	X	x	o	o	O	o	O	o	
D1	X	x	x	X			o	O			o	
B1	o							o				
D2	X	X					x		o	o		
D4	x	O	o				x	X			o	o
B4		x								x		x
D5	o	o	O			x	o	X		x	x	
B5						x	x			x		o
D6		o	O	o	x		x	x				x
B7		o	X			o					X	x
D7	O	O		X	o			x				o

- a patch of positive residuals in the upper left corner, corresponding to high Democratic votes in the westernmost of our 11 counties in 1920 up toward 1936.

- negative residuals for the southeasternmost county, D7, in 1920 and 1924, followed by a positive one in 1932.

- positive residuals in the lower right corner, corresponding to high democratic votes in the eastern counties after 1940.

Clearly there are strong indications of trends both with time and from East to West. The latter seem heaviest from 1920 to 1928 and in 1948.

The resistant analysis, particularly as displayed by panel C, is clearly very much better, both in itself and as a first step.

With two sets of trends already apparent, we are ready for more than one step of peeling off layer after layer of appearances. It is up to the interested reader to see just how many layers there are for this example. As many as for the earlier Nebraska example (chapter 17)? One less? One or two more? Whichever the answer, our main point is clearly made: the need to strip layer after layer was neither the purpose nor the result of our gerrymandering.

19D. Relax, relax

* stopping sooner, stripe by stripe *

We have so far seemed to take the view that when we found a stripe median not zero we should do something about it. Panel A of exhibit 10 begins in the same way as panel A of exhibit 5, finding column medians for the residuals of exhibit 3.

For the first column, the median is 0. For the second column, the median is 15. At first glance 15 looks large -- after all only two other medians out of 12 are larger. When we look back at exhibit 3, however, we find that the three middle values are -28, 15, and 26. With such broad spacings as this, we put a lot of weight on our single value if we adjust this whole column by -15. Should we bother to adjust at all? Perhaps not.

Going over the spacings in each stripe is extra work. We are looking for less work, not more. The simplest thing to look at is our count of "-", "0" and "+" entries that we have used in finding the median itself. For this column, these counts were 5, 0, 6. We might as well plan to not bother with such cases.

Which cases? For 11 items per stripe, where the ideal pattern is 5, 1, 5 (five negative, one zero, 5 positive) we ought not to bother unless the - or + count is at least 2 larger than ideal. The same for all odd numbers of entries.

exhibit 10 of chapter 19 -- Nebraska voting 2

Pomedian polishing the residuals of exhibit 3
-- all in 0.1% Democratic

A) POMEDIANS for COLUMNS -- 5, 1, 5 ideal; values mostly repeated from exhibit 5

Election	'20		'28		'36		'44		'52		'60	
Count -	5	5	8	8	6	4	5	6	3	3	6	6
Count 0	1	0	0	0	0	0	0	1	1	0	0	0
Count +	5	6	3	3	5	7	6	4	7	8	5	5
Median	0	15	-28	-28	-3	5	2	-14	3	11	-3	-8
Pomedian	0	0	-28	-28	0	5	0	-14	3	11	0	0

B) EFFECT ANALYSIS of exhibit 3 ADJUSTED for COLUMN POMEDIANS

													eff
DO	15	61	284	117	29	-44	-39	-103	-66	-125	-49	-22	37
D1	91	61	39	135	-13	-36	-40	-77	0	-16	-41	-22	-69
B1	-29	26	21	0	16	9	9	-8	0	-3	15	-8	16
D2	62	96	-19	-13	3	-25	25	14	-38	-48	-13	2	16
D4	26	-49	-19	8	17	20	42	78	-3	-6	-41	-22	15
B4	-13	47	-4	-18	-12	-2	-6	-13	16	20	-3	45	-16
D5	-24	-34	-69	-1	-3	26	-26	118	8	34	26	4	-49
B5	11	15	1	6	-11	22	35	-1	3	19	-17	-29	-42
D6	0	-28	-67	-20	30	0	2	41	13	0	14	43	-36
B7	-30	-54	135	7	-16	-39	-19	0	-16	1	71	30	-55
D7	-104	-142	9	44	-25	18	22	77	42	31	39	-15	-73
eff	-48	-89	-81	254	129	23	-2	54	-133	-94	-101	32	349

(cont'd.)

exhibit 10 of chapter 19 (cont'd.)

C) POMEDIANS for ROWS of panel B -- 5, 1, 1, 5 ideal

County	D0	D1	B1	D2	D4	B4	D5	B5	D6	B7	D7
Count -	7	7	4	6	6	8	6	4	3	6	4
Count 0	0	1	2	0	0	0	0	1	3	1	0
Count +	5	4	6	6	6	4	6	7	6	5	8
(m	(-22	(-13	(0	(-13	(-3	(-3	(-1	(1	(0	(-16✓	(18
(m	(-39✓	(-16✓	(9	(2	(8	(-4✓	(4	(3	(2	(0	(22✓
Median	-30	-14	4	-6	2	-4	2	2	1	-8	20
Pomedian	-30	-14	0	0	0	-4	0	2	0	0	20

D) EFFECT ANALYSIS of panel B adjusted for panel C

	'20		'28		'36		'44		'52		'60		eff
D0	45	91	314	147	59	-14	-9	-73	-36	-95	-19	8	7
D1	105	75	53	149	1	-22	-26	-63	14	-2	-27	-8	-83
B1	-29	26	21	0	16	9	9	-8	0	-3	14	-8	16
D2	62	96	-19	-13	3	-25	-25	14	-38	-48	-13	2	16
D4	26	-49	-19	8	17	20	42	78	-3	-6	-41	-22	15
B4	-9	51	-	-14	-8	2	-2	-9	20	24	1	49	-20
D5	-24	-34	-69	-1	-3	26	-26	118	8	34	26	4	-49
B5	9	13	-1	-2	-13	20	33	-3	1	17	-19	-31	-40
D6	0	-28	-67	-20	30	0	2	41	13	0	14	43	-36
B7	-30	-54	135	7	-16	-39	-19	0	-16	1	71	30	-55
D7	-124	-162	-11	24	-45	-2	2	55	22	11	19	-35	-53
eff	-48	-89	-81	254	129	23	-2	54	-133	-94	101	32	349

(cont'd.)

exhibit 10 of chapter 19 (cont'd.)

E) POMEDIANS for COLUMNS of panel D -- 5, 1, 5 ideal

Election	'20		'28		'36		'44		'52		'60	
Count -	5	5	6	5	5	5	6	5	4	5	5	5
Count 0	1	0	1	1	0	1	0	0	1	1	0	0
Count +	5	6	4	5	6	5	5	5	6	5	6	6
Median	0	13	-1	0	1	0	-2	0	1	0	1	2
Pomedian	0	0	0	0	0	0	0	0	0	0	0	0

F) STEM-and-LEAF for the RESIDUALS of panel D

High	Low	To		
105	-124	8	H	see column
91	-162	11	7	581
96	-95	12	6	2
314	(-73)	16	5	1395
135	(-69)	21	4	52193
147	(-63)	25	3	0340
149		36	2	66140600246
118		47	1	36744371549
(78)		(22)	0	900871392922008101182 4
(75		63	-0	9011283029289332368 8
(71)		43	-1	9913435496939
(62)		30	-2	94802565672
		19	-3	0496815
		12	-4	9581
		8	-5	4
		7	-6	973
		4	-7	3
		3	L	see column

47 + 22 + 63 = 132 ✓ hinges at 33h

Hinges:	+20h	and -19
Midspread:		39h
Sides:	+60	and -58h
Outside:	12	and 7
(count/20):		6h

(cont'd.)

exhibit 10 of chapter 19 (cont'd.)

G) The ESSENTIAL PARTS of panel A -- up to 6

Election	'20		'28		'36		'44		'52		'60	
Count -	5	5	8	8	6	4	5	6	3	3	6	6
Count 0	1	0	0	0	0	0	0	1	1	0	0	0
Count +	5	6	3	3	5	7	6	4	7	8	5	5
Median			-28	-28		5		-14	3	11		
Pomedian	0	0	-28	-28	0	5	0	-14	3	11	0	0

H) The ESSENTIAL PARTS of panel C -- up to 6

County	D0	D1	B1	D2	D4	B4	D5	B5	D6	B7	D7
Count -	7	7	4	6	6	8	6	4	3	6	4
Count 0	0	1	2	0	0	0	0	1	3	1	0
Count +	5	4	6	6	6	4	6	7	6	5	8
m	-22	-13				-3					18
m	-39	-16				-4		3			22
Median	-30	-14				-4		2			20
Pomedian	-30	-14	0	0	0	-4	0	2	0	0	20

I) The ESSENTIAL PARTS of panel E -- up to 6

Election	'20		'28		'36		'44		'52		'60	
Count -	5	5	6	5	5	5	6	5	4	5	5	5
Count 0	1	0	1	1	0	1	0	0	1	1	0	0
Count +	5	6	4	5	6	5	5	5	6	5	6	6
Median												
Pomedian	0	0	0	0	0	0	0	0	0	0	0	0

J) PROBLEMS

1) Check the arithmetic by making a difference analysis from exhibit 3 to panel D above.

2) Compare with the analysis of exhibit 6 by making a difference analysis.

For 12 items per stripe -- which we face in this example when we seek row medians -- the ideal pattern -- when we prefer to separate the two residuals that contribute to the median -- is 5, 1, 1, 5, where the two 1's combine to fix the median. Again we ought to allow some fairly common patterns to call for no change. Here, as panel C (of exhibit 10) illustrates, it is enough to allow one more than 5 (5 not 6) either positive or negative. We can do the same for all batches with an even count.

The number we are now ready to use -- the median if the batch is sufficiently unbalanced in sign, zero otherwise -- deserves a name. To recognize its origin -- a polishing tool -- we call it the

pomedian

|| Its great virtue is that it saves work. ||

Panels G, H, and I (of exhibit 10) show the minor way in which it saves work. They repeat the parts of panels A, C and E that we need if we are sure we are going to use the pomedian. Notice that we save half the calculations of medians in the first two cases, all in the third.

Saving calculation of the median is nice. It is nicer, and more important, however, that when a median or pomedian is zero, we can adjust for it by just copying that stripe without change. This saves much more work.

How does this analysis compare with that of exhibit 6? We haven't checked exhibit 9 yet, so we don't know about errors. This aside, though, the gross behavior is surely similar. We have

	Midspread	Outside
exhibit 6 (uncorrected)	40	12 and 9
exhibit 9	39h	12 and 7

Presumably the analysis of exhibit 9 is quite acceptable -- and took much less arithmetic.

* judgment often softens further *

Our introduction of pomedians has two consequences, not just one. It is important that it offers us an easier routinized calculation. It is also important that it emphasizes that we do not have to go all the way to zero medians.

Pomedian polishing is not expected to replace the use of judgment in choosing when to stop. It should encourage the use of judgment, especially in stopping before all pomedians are zero.

* checking should be relaxed *

Our repeated call for checking of possibility -- followed by making the analysis possible and going on is also a call for us to relax wherever we can. We dare not relax about

whether our analysis analyzes our data, but we can and do relax about just how it came to be. Checking has to be there, but it should be made easy and used in a relaxed way.

19E. Tables with holes

Now that we are well on our way toward a light-hearted attitude toward two-way fitting, it is time to look at tables which do not have all the entries ready for analysis -- either because the entries never were, or they never reached us, or we decided it was better to set them aside.

* after an analysis *

At the close of chapter 18, we recognized that median polishing, while greatly improving our analysis of the ratio of dry weight to wet weight for wheat grain, did not deal with the one very anomalous value as well as we would have liked. While the size of the wild value's residual does not matter, its sign does. As the reader may easily verify, changing the original value from 1035 (1000ths of wet weight) to 1135 or 935 will not change the fit by median polishing, while changing to 335,435 or 535 will.

The easy way to avoid the ill effects of the wild value is to declare it "absent". Now we can polish anew. Exhibit 11 shows the calculations, carried further

exhibit 11 of chapter 19 -- dry/wet ratios

Median polishing with an absent value

A) The INITIAL ANALYSIS -- with medians of residuals, from exhibit 34 of chapter 18.

	Nitrogen applied					
Block	None	Early	Middle	Late	eff	(med)
1	-1	-38	1	25	-3	(0)
2	5	10	-9	-52	-2	(0)
3	-26	abs	19	-20	8	(-20)
4	1	-1	-1	8	3	(0)
eff	-31	20	-17	7	753	
(med)	(0)	(-11)	(0)	(-6)	(0)	

B) POLISHED by COLUMN MEDIANS

1	-1	-27	1	31	-3	(0)
2	5	21	-9	-46	-2	(-2)
3	-26	abs	19	-14	8	(-14)
4	1	0	-1	14	3	(0)
eff	-31	20	-17	7	753	(0)

C) REPOLISHED by ROW MEDIANS

1	-1	-27	1	31	-3	
2	7	23	-7	-44	-4	
3	-12	abs	33	0	-6	
4	1	0	-1	14	3	
eff	-31	20	-17	7	753	
(med)	(0)	(0)	(0)	(7)	(-4)	

D) AGAIN REPOLISHED -- by COLUMN MEDIANS

1	-1	-27	1	24	1	(0)
2	7	23	-7	-51	0	(0)
3	-13	abs	33	-7	-10	(-7)
4	1	0	-1	7	-1	(0)
eff	-31	20	-17	14	753	(-2)

(cont'd.)

exhibit 11 (cont'd.)

E) YET AGAIN REPOLISHED -- by row median

1	-1	-27	1	24	1
2	7	23	-7	-51	0
3	-6	abs	40	0	-3
4	1	0	-1	3	-1
eff	-29	22	-15	16	751
(med)	(0)	(0)	(0)	(2)	(0)

F) ONCE AGAIN ...

1	-1	-27	1	22	1	(0)
2	7	23	-7	-53	0	(0)
3	-6	abs	40	-2	-3	(-2)
4	1	0	-1	1	1	(0)
eff	-29	22	-15	16	751	(0)

G) AND ONCE MORE

						fit
1	-1	-27	1	22	1	756
2	7	23	-7	-53	0	755
3	-6	abs	40	0	-1	754
4	1	0	-1	1	1	755
eff	-33	18	-19	22	755	
fit	722	773	736	777		
(med)	(0)	(0)	(0)	(0)		

than needed in order to show how much declaring this one value absent pulls together the results -- both fit and residuals. Exhibit 12 lets us compare by eye residuals, block fits, and treatment fits before and after. The pulling together of the first two is quite clear.

* from the beginning *

It can be interesting to take this same example and begin at the beginning, leaving out the impossible value before we start to analyze the table. Let us, at the same time, modify our polishing technique, taking row medians of only those entries belonging to complete columns and, similarly, taking column medians of only entries belonging to complete rows. Usually it pays to do this for at most 2 half-steps, especially since we would like the unmodified medians small when we stop. Exhibit 13 shows the computation.

Notice the check, correct, and do not go back process in panels F and G.

When, as we have done in panel F, we use

original - final

as our checking difference analysis, we get the detailed fit brought to our attention, since it appears as the residuals block. (The negatives of its parts also appear in the effect stripes and in common.) This bonus makes such an end-to-end difference analysis a specially useful form of check.

exhibit 12 of chapter 19 -- dry/wet ratios in 1000ths

Effect of median polishing after declaring one value absent
(before = exhibit 34 of chapter 18; after = exhibit 1)

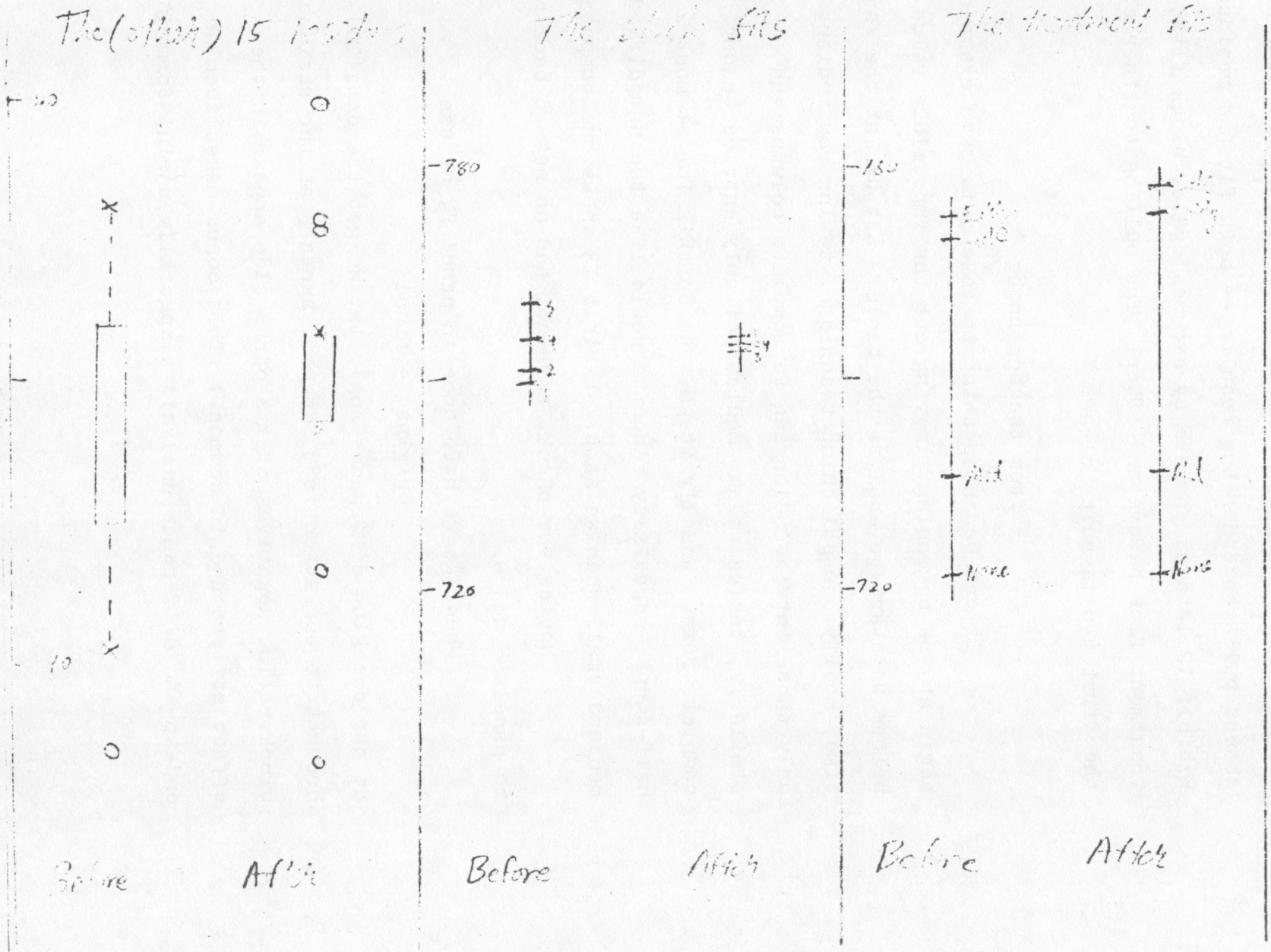

exhibit 12 of chapter 19 -- dry/wet ratios

exhibit 13 of chapter 19 -- dry/wet ratios

Modified median polishing from the beginning for a table with a hole

) The DATA -- in 1000ths

	None	Early	Mid	Late	(134) (med)*	(med)
1	718	732	734	792	(734)	(733)
2	725	781	725	716	(725)	(725)
3	764	abs	763	758	(758)	(758)
4	726	765	738	781	(738)	(752)

* 134 med - median of columns 1, 3, and 4.

) ONCE adjusted -- for 134 median

1	-16	-2	0	58	734
2	0	56	0	-9	725
3	-54	abs	5	0	758
4	-12	27	0	43	738
(124 med)	(0)	(27)	(0)	(43)	(738)
(med)	(-6)	(27)	(0)	(22)	

124 med = median of rows 1, 2, and 4.

) TWICE adjusted -- for 124 median

1	-16	-29	0	15	-4	(0)	(-8)
2	0	29	0	-52	-13	(0)	(0)
3	-54	abs	5	-43	20	(-43)	(-43)
4	-12	0	0	0	0	(0)	(0)
ef							
eff	0	27	0	43	738	(0)	(14)

) THREE TIMES adjusted -- now for row medians

	-8	-21	8	23	4	
	0	29	0	-52	-13	
	-11	abs	48	0	-23	
	-12	0	0	0	0	
	-14	13	-14	29	752	
(124 med)	(-11)	(0)	(0)	(0)	(0)	
(med)	(-10)	(0)	(4)	(0)	(-6)	

exhibit 13 (cont'd.)

E) FOUR TIMES adjusted -- now for column medians

					(med)
2	-21	4	23	10	(3)
10	29	-4	-52	-7	(3)
-1	abs	44	0	-17	(0)
-2	0	-4	0	6	(-1)
-14	13	-14	29	746	

F) CHECK -- panel A MINUS panel E = fit (both signs*)

					check sum
716	753	730	769	-10	759
✓	✓	✓	✓		
715	752	729	768	7	775✓
✓		✓	✓		
705	abs	719	758	17	775✓
✓		✓	✓		
728	765	742	781	-6	775✓
14	-13	14	-29	-746	-775✓
742	752✓	756	752✓	-752✓	

*The entries in the residual block are the fit; the entries in effect stripes and in common are the negatives of the parts of the fit.

G) CORRECTED VERSION of panel E -- (don't compare with panel F)

					eff	(eff^+)
	2	-21	4	23	-6	(0)
	10	29	-4	-52	-7	(-1)
	-1	abs	44	0	-17	(-11)
	-2	0	4	0	6	(12)
eff	-24	13	-10	29	746	
(eff^+)	(-26)	(11)	(-12)	(27)		(742)

(eff^+) stripes so eff stripes altered to bring effect medians to zero.

19F. More holes may make more trouble

All too often, we have to deal with tables with more than a few holes. Sometimes these give no particular trouble -- and sometimes quite the contrary.

Exhibit 14 uses a portion of the data on the yield of straw under eight manurial treatments given by Eden and Fisher in a classical paper. (We will come back to this data more seriously in the next chapter.) We have selected 5/8ths of the original data in a simple pattern as a basis for showing what can happen. For this set of data, 37.5% of which is holes, we find that three half-steps of "plain old median polish" makes all stripe medians zero, producing an analysis that seems quite reasonable. In this case, we seem to have done all right.

A doubting Thomas, after he looks over exhibit 14 quite carefully, is likely to say:

> "This was an easy test. The treatments with the most different fits are E and F. The observed yields of straw for E varied from 232 to 393 while those for F ran from 200 to 376, clearly a lot of overlap. What would happen if the treatments were very different?"

The simple response to this heckle is to modify this data, say by adding 1000 to the treatment A entries, 200 to those for treatment B, and so on, thus making some substantial treatment effects. Exhibit 15 analyzes the modified data.

exhibit 14 of chapter 19 -- straw yield, 5/8ths

Use of plain median polishing on a table with 37.5% holes

A) The DATA

Block	Treatment A	B	C	D	E	F	G	H
1	242	a	a	a	322	200	260	203
2	321	382	a	a	a	261	318	275
3	261	201	298	a	a	a	266	207
4	317	316	381	255	a	a	a	331
5	255	280	300	238	232	a	a	a
6	a	285	294	309	393	258	a	a
7	a	a	256	283	351	306	276	a
8	a	a	a	324	363	376	385	328

B) ONCE adjusted -- for row medians (column 1 is an apparent accident)

									eff
	0	a	a	a	80	-42	18	-39	242
	0	61	a	a	a	-60	-3	-46	321
	0	-60	37	a	a	a	5	-54	261
	0	-1	64	-62	a	a	a	14	317'
	0	25	45	-17	-23	a	a	a	255
	a	-9	0	15	99	-36	a	a	294
	a	a	-27	0	68	12	-18	a	283
	a	a	a	-39	0	13	22	-35	363
(med)	(0)	(-1)	(37)	(-17)	(68)	(-36)	(5)	(-39)	(288)

C) TWICE adjusted -- this time for column medians

										(med)
	0	a	a	a	12	-6	13	0	-46	(0)
	0	62	a	a	a	-24	-8	-7	33	(-7)
	0	-59	0	a	a	a	0	-15	-27	(0)
	0	0	27	-45	a	a	a	53	29	(0)
	0	26	8	0	-91	a	a	a	-33	(0)
	a	-8	-37	32	31	0	a	a	6	(0)
	a	a	-64	17	0	48	-23	a	-5	(0))
	a	a	a	-22	-68	49	17	4	75	(4)
	0	-1	37	-17	68	-36	5	-39	288	(2)

exhibit 14 (cont'd.)

D) THREE times adjusted

	0	a	a	a	12	-6	13	0	-46
	7	69	a	a	a	-17	-1	0	26
	0	-59	0	a	a	a	0	-15	-27
	0	0	27	-45	a	a	a	53	29
	0	26	8	0	-91	a	a	a	-33
	a	-8	-37	32	31	0	a	a	6
	a	a	-64	17	0	48	-23	a	-5
	a	a	a	-20	-72	45	13	0	79
	-2	-3	35	-19	66	-38	3	-37	290
(med)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

E) PROBLEMS

1) Make a stem-and-leaf display of the residuals, and show that the clarity is between 8 and 9 when the midspread is calculated normally, but between 4 and 5 if all zero residuals are set aside before calculating the midspread.

2) Check the possibility of the analysis.

F) SOURCE

T. Eden and R. A. Fisher 1927. Studies in Crop Variation. IV. The experimental determination of the value of top dressings with cereals. Journal of Agricultural Science 17: 548-562.

exhibit 15 of chapter 19 -- 5/8ths modified

The same analysis applied to modified data

A) The MODIFIED DATA

	A	B	C	D	E	F	G	H		
1	1242	abs	abs	abs	5322	6200	7260	8203	0	(6200)
2	1321	2382	abs	abs	abs	6261	7318	8275	0	(6261)
3	1261	2201	3298	abs	abs	abs	7366	8207	0	(3298)
4	1317	2316	3381	4255	abs	abs	abs	8331	0	(3381)
5	1255	2280	3300	4238	5232	abs	abs	abs	0	(3300)
6	abs	2285	3294	4309	5893	6258	abs	abs	0	(4309)
7	abs	abs	3256	4283	5351	6306	7276	abs	0	(5351)
8	abs	abs	abs	4324	5363	6378	7385	8328	0	(6376)

B) ONCE adjusted -- by row medians

1	-4958	abs	abs	abs	-978	0	1060	2003	6200
2	-4940	-3879	abs	abs	abs	0	1057	2014	6261
3	-2037	-1097	0	abs	abs	abs	4068	4909	3298
4	-2064	-1065	0	874	abs	abs	abs	4950	3381
5	-2045	-1020	0	938	1932	abs	abs	abs	3300
6	abs	-2024	-1015	0	1084	1949	abs	abs	4309
7	abs	abs	-2095	-1068	0	955	1925	abs	5351
8	abs	abs	abs	-2052	-1013	0	1009	1932	6376
	0	0	0	0	0	0	0	0	0
(med)	(-2064)	(-1097)	(0)	(0)	(0)	(0)	(1060)	(2014)	(4830)

(cont'd.)

exhibit 15 (cont'd.)

C) TWICE adjusted -- this time column medians

										(med)
1	-2892	abs	abs	abs	-978	0	0	-11	1370	(-11)
2	-2876	-2872	abs	abs	abs	0	-3	0	1431	(-3)
3	27	0	0	abs	abs	abs	3008	2895	-1532	(27)
4	0	32	0	874	abs	abs	abs	2936	-1449	(32)
5	19	77	0	938	1932	abs	abs	abs	-1530	(77)
6	abs	-927	-1015	0	1084	1949	abs	abs	-530	(0)
7	abs	abs	-2095	-1068	0	955	865	abs	521	(0)
8	abs	abs	abs	-2052	-1013	0	-51	-62	1546	(-62)
	-2064	-1097	0	0	0	0	1060	2014	4830	(0)

D) THREE TIMES adjusted -- this time row medians

1	-2891	a	a	a	-967	11	11	0	1381
2	-2873	-2774	a	a	a	3	0	3	1428
3	0	-27	-27	a	a	a	2981	2867	-1505
4	-32	0	-32	842	a	a	a	2904	-1417
5	-58	0	-77	861	1855	a	a	a	-1453
6	a	-927	-1015	0	1084	1949	a	a	-530
7	a	a	-2095	-1068	0	955	865	a	521
8	a	a	a	-1990	-951	62	11	0	1484
	-2064	-1097	0	0	0	0	1060	2014	4830
(med)	(-58)	(-27)	(-77)	(0)	(0)	(62)	(11)	(3)	(-4)

(cont'd.)

exhibit 15 (cont'd.)

E) FOUR TIMES adjusted

										(med)
1	-2923	a	a	a	-967	-51	0	-3	1385	(-51)
2	-2815	-2752	a	a	a	-59	-11	0	1432	(-59)
3	58	0	50	a	a	a	2970	2864	-1501	(58)
4	26	27	45	842	a	a	a	2901	-1413	(45)
5	0	27	0	861	1855	a	a	a	-1449	(27)
6	a	-900	-938	0	1084	1887	a	a	-526	(0)
7	a	a	-2018	-1068	0	893	854	a	525	(0)
8	a	a	a	-1990	-951	0	0	-3	1488	(-3)
	-2122	-1124	-77	0	0	62	1071	2017	4826	(0)

After four half-steps the analysis is quite unsatisfactory, however you look at it:

- The sums of sizes of medians after steps 2, 3 and 4 are 208, 237, and 243 -- hardly rapidly decreasing. We can clearly expect many more half steps.

- The analysis itself is clearly far from adequate. The differences of the form "residual for treatment E minus residual for treatment D" are 994, 1084, 1068 and 1039 (in blocks 5, 6, 7, 8 respectively). Clearly there is a large difference between these treatments yet unfitted.

- The residuals here CAN be made as small as those of panel G of exhibit 14, of which the largest in size was -92. Yet we have five residuals of size greater than 2800.

It would be worse, of course, if we had some, but not all of these troubles. To get rid of the unnecessarily large residuals, we need to change the stripe effects by a total of about 20,000. If we get a change of 200 in the right direction each time, we may get there in only 100 half-steps. If the half-steps were rapidly shrinking in size, we would know that we could never get there. (How far is 100 half-steps from never?)

To analyze data such as these, we need a better approach.

* a clearer example

Hard looking at this example shows us how to make a simpler one. Exhibit 16 has entries between -3 and +3 with the same pattern of absent values. All stripe medians vanish, though as panel B shows, there can be an analysis with all residuals zero. For such holey tables, there must be another approach.

Clearly panels A and B (of exhibit 16) show that there can be more than one analysis of the same data with all stripe medians zero. This example involves many absent values. There can be examples for complete tables, however, as some of us will see when and if section 19K is reached.

19G. Differences can help

In which direction ought we to seek for a better analysis of exhibit 16? We want resistance and we do not want to be stalled at panel A of that exhibit. How did we decide that exhibit 15 was in trouble? By looking at differences -- there between columns and within rows -- and finding them consistent. It is thus natural to use differences to find a better fit.

* adjacent columns *

Exhibit 17 collects differences of this sort for adjacent columns in panel E of exhibit 15, and then cumulates them to indicate a possible collection of additional column fits. Clearly these differences are telling us something.

exhibit 16 of chapter 19 -- casual example

How median polish may not work

A) The DATA -- all stripe medians zero

	A	B	C	D	E	F	G	H	(med)
1	-3	a	a	a	-1	0	0	0	(0)
2	-3	-3	a	a	a	0	0	0	(0)
3	0	0	0	a	a	az	3	3	(0)
4	0	0	0	1	a	a	a	3	(0)
5	0	0	0	1	2	a	a	a	(0)
6	a	-1	-1	0	1	2	a	a	(0)
7	a	a	-2	-1	0	1	1	a	(0)
8	a	a	a	-2	-1	0	0	0	(0)
(med)	a	a	a	0	0	0	0	0	(0)

B) A SENSIBLE ANALYSIS of the DATA

1	0	a	a	a	0	0	0	0	-2
2	0	0	a	a	a	0	0	0	-2
3	0	0	0	a	a	a	0	0	1
4	0	0	0	0	a	a	a	0	1
5	0	0	0	0	0	a	a	a	1
6	a	0	0	0	0	0	a	a	0
7	a	a	0	0	0	0	0	a	-1
8	a	a	a	0	0	0	0	0	-2
eff	-1	-1	-1	0	1	2	2	2	0

exhibit 17 of chapter 19 -- modified 5/8ths

Differences within row for panel E of exhibit 15

	B minus A	C minus B	D minus C	E minus D	F minus E	G minus F	H minus G
1	a	a	a	a	916	51	-3
2	63	a	a	a	a	48	11
3	-58	50	a	a	a	a	-106
4	1	18	797	a	a	a	a
5	27	-27	861	994	a	a	a
6	a	-38	938	1084	803	a	a
7	a	a	950	1068	893	-39	a
8	a	a	a	1039	951	0	-3
(med)	(14)	(-4)	(900)	(1054)	(904)	(24)	(-3)

	A	B	C	D	E	F	G	H
(cum*)	0	14	10	910	1964	2868	2892	2889

* cum = cumulation of differences, starting at 0 for convenience.

It is of passing interest to put these new column fits to use. Exhibit 18 does this, carrying on with the obvious further half-steps. After the conventional 3 more half-steps, the stripe medians are between 1 and 0, and the analysis looks quite reasonable.

* other choices *

If there had been no absent values, looking at adjacent differences is not the easy thing to do. It is easier to use differences of all other columns with a single column, even if we do not short cut. The process of cumulation cumulates irregularities, as well as what we want. Exhibit 19 shows how we can make such calculations -- and how we can short-cut them.

The short-cut is a technique we are wise to learn. Picking out a good column or row to adjust for can help a lot when we would otherwise be in trouble.

An energetic man would perhaps like to deal with the differences between each possible pair of columns. If you want to do this, you face a further difficulty. We already found medians for

column B minus column A:	14
column C minus column B:	-4
column C minus column A:	0

These three differences cannot come from any one set of additional fitted values. To use all differences we must do something about inconsistency.

exhibit 18 of chapter 19 -- modified 5/8ths

Use of the additional fit of exhibit 17 to finish fitting the data of exhibit 15

A) ONCE MORE adjusted -- the analysis of panel E of exhibit 15 adjusted for the cumulated differences of exhibit 17

	A	B	C	D	E	F	G	H	eff	(med)
1	-2823	a	a	a	-2903	-2919	-2892	-2892	1385	(-2892)
2	-2815	-2766	a	a	a	-2927	-2903	-2889	1432	(-2889)
3	58	-14	40	a	a	a	-78	-25	-1501	(40)
4	26	13	35	-68	a	a	a	12	-1413	(13)
5	0	13	-10	-49	-109	a	a	a	-1449	(-10)
6	a	-914	-948	-910	-880	-981	a	a	-526	(-914)
7	a	a	-2018	-1978	-1964	-2025	-2038	a	525	(-2018)
8	a	a	a	-2900	-2915	-2868	-2892	-2892	1488	(-2892)
	-2122	-1110	-67	-910	1964	2930	3963	4906	4826	(1437)

B) TWICE MORE adjusted -- panel A adjusted for its row medians

1	69	a	a	a	-11	-27	0	0	-1507
2	74	123	a	a	a	-38	-14	0	-1457
3	18	-54	0	a	a	a	38	-65	-1461
4	13	0	22	-81	a	a	a	-1	-1400
5	10	23	0	-39	-99	a	a	a	-1459
6	a	0	-34	4	34	-67	a	a	-1440
7	a	a	0	40	54	-7	-20	a	-1493
8	a	a	a	-8	-23	30	0	0	-1404
eff	-3559	-2547	-1514	-527	527	1493	2526	3469	6263
(med)	(18)	(0)	(0)	(-8)	(-11)	(-27)	(0)	(0)	(-1458)

(cont'd.)

exhibit 18 (cont'd.)

C) THRICE MORE adjusted -- panel B adjusted for column medians of its residuals

									eff	(med)
1	51	a	a	a	0	0	0	0	-49	(0)
2	56	123	a	a	a	-18	-14	0	1	(0)
3	0	-54	0	a	a	a	38	-65	-3	(0)
4	-5	0	22	-73	a	a	a	-1	58	(-1)
5	-8	23	0	-31	-88	a	a	a	-1	(-8)
6	a	0	-34	12	45	-40	a	a	18	(0)
7	a	a	0	48	65	20	-20	a	-35	(20)
8	a	a	a	0	-12	57	0	0	54	(0)
eff	-3541	-2547	-1574	-535	516	1466	2526	3469	4805	(-10)

D) FOUR TIMES MORE adjusted -- panel C adjusted for row medians of its residuals

1	51	a	a	a	0	0	0	0	-49
2	56	123	a	a	a	-18	-14	0	1
3	0	-54	0	a	a	a	38	-65	-3
4	-4	1	23	-72	a	a	a	0	57
5	0	31	8	-23	-80	a	a	a	-9
6	a	0	-34	12	45	-40	a	a	18
7	a	a	-20	28	45	0	-40	a	-15
8	a	a	a	0	-12	57	0	0	54
eff	-3531	-2537	-1564	-525	525	1476	2536	3479	4795
(med)	(0)	(1)	(0)	(0)	(0)	(0)	(0)	(0)	(-1)

E) PROBLEMS

1. Check the possibility of this analysis, and correct it where necessary.

2. Apply the procedure used here -- one half-step by cumulated medians of adjacent differences, then three half-steps by median polishing.

exhibit 19 of chapter 19 -- 5/8ths modified

Using a single column as a basis for adjusting an incomplete table

A) ANOTHER set of DIFFERENCES between COLUMNS of panel E of exhibit 15.

	B minus A	C minus A	D minus A	E minus A	F minus A	G minusq A	H minus A
1	a	a	a	1856	2772	2823	2820
2	63	a	a	a	2756	2804	2815
3	-58	-8	a	a	a	2912	2806
4	1	19	816	a	a	a	2875
5	27	0	861	1855	a	a	a
6	a	a	a	a	a	a	a
7	a	a	a	a	a	a	a
8	a	a	a	a	a	a	a
(med)	(14)	(0)	(838)	(1856)	(2764)	(2823)	(2818)

A	B	C	D	E	F	G	H
0	14	0	838	1856	2764	2828	2818

Example: 0 arbitrary, 14 = 0 + 14, 0 = 0 + 0, etc.

B) The result of adjusting panel E of exhibit 15 by the ENTRIES of its FIRST COLUMN

1	0	a	a	a	1856	2772	2823	2820
2	0	63	a	a	a	2756	2804	2815
3	0	-58	-8	a	a	a	2912	2806
4	0	1	19	816	a	a	a	2875
5	0	27	0	861	1855	a	a	a
6	a	A	A	A	A	A	A	A
7	a	A	A	A	A	A	A	A
8	a	A	A	A	A	A	A	A
(med)	(0)	(14)	(0)	(838)	(1856)	(2764)	(2823)	(2818)

a = originally absent, A = adjusting value absent

C) PROBLEMS

1) Complete the analysis here began.

2) Check the possibility of your final analysis.

****

19H. How far have we come?

This chapter has had four aims:

- to REMOVE any thought that Nebraska politics can be well analyzed by procedures that are NOT resistant.

- to show that resistant procedures need not be more difficult to carry through by hand than those that are not resistant.

- to RELAX any rigid attitudes that seem to call for "THE analysis" instead of for "a reasonable analysis".

- to show that tables with few holes need cause little difficulty, but that some tables with many holes need care.

Our new example of 11 counties in northeast Nebraska proved to be quite clear and definite. Assorted rather deviant values have undesirably large effects on an analysis by means. Indeed we find that the changes in residuals by means due to polishing are about half as large as the residuals after polishing. This is rather too much for comfort, even in an exploratory situation.

This example offered a chance to illustrate checking -- and then going right on -- on a larger table.

Once we had a possible, resistant analysis, we saw that there were many layers to be stripped away by further analysis.

We tried to teach ourselves to relax wherever we could in making two-way analyses. For those who need a rule, there is the pomedian, whose value is either that of the median or zero. The pomedian vanishes when the residuals are nearly balanced in sign. Its use calls for fewer half-steps of polishing. The use of judgment to still further reduce the number of half-steps is encouraged.

"Nearly balanced" may be taken to be exactly this:

- suppose the median zero and all values different, and set aside the one or two observations that contribute to the median -- equal counts of positive and negative values will remain.
- increase this count by one.
- if neither the count of positive (not zero) values in the actual batch or the-count of negative (not zero) values exceeds this increased count, call the situation "close to" and set the pomedian to zero.

Tables with holes can be approached in many ways, choosing an adjustment for any particular stripe:

- by taking medians of all entries.
- by taking medians of only those values from complete stripes of the other kind.
- by adjusting for the entries in one particular stripe of the other kind.

We expect to finish our polishing using the first of these -- medians of all entries in a stripe. We can often gain from using one of the others once or twice.

A table with 37.5% holes, and all stripe medians zero can demand analysis so complete as to leave zero residuals. In one such, adjusting for the entries of one stripe puts us safely on our way.

The key to dealing with seriously holey tables is to work with differences. Adjusting for the entries from one stripe is just a special case.

It is natural to try to use all differences, but this is easily seen to require a little more thought.

****

19K. <u>Uniqueness?</u> (Optional background)

* first, means *

To require that a two-way analysis has all its stripe MEANS zero fixes the analysis uniquely.

This is a consequence of the important fact that

difference of means ≡ mean of differences.

If we had two different analyses of the same data, both with stripe means zero, the difference analysis has to be very special. Consider any stripe. In each of the original analyses the mean of its entries is zero. In the difference analysis we have the differences of each pair of corresponding entries. Their mean -- as the mean of differences identically equal to the difference of means -- must vanish. Consider any two parallel stripes, the means of their entries are both zero, differing by zero, hence the corresponding stripe effects must be equal. Now we know that all row effects are the same -- and that all column effects are the same. We can modify each analysis, without altering the residuals, to make all row effects zero and all column effects zero. We merely put whatever is needed into the common term. It is easy to see that once this is done, each common term is also zero. Since the difference analysis analyzes zero, and its effects

and common term are zero, all its residuals are zero. Hence the residuals in the two analyses are exactly the same.

|| Two analyses with all stripe means zero have the same residuals. ||

* now, medians *

|| Nothing of the sort need be true for MEDIANS. ||

Exhibit 20 repeats the upper left corner of the data of the example of chapter 17, and gives four different analyses for this 3×4 table that have all stripe medians zero.

The first two of these manage to make the H3 median vanish by rounding 1/2 to zero. They both do this in the same way, and a change from 278 to 279 in data value would make them a perfect example of non-uniqueness.

The third and fourth are a perfect example of non-uniqueness. If you look hard at them (perhaps making a difference analysis) you can see how to construct still more analyses of this same data with all stripe medians zero.

* the morals *

What morals ought we to draw? One is easy: If we want to specify a unique two-way analysis, and stripe medians are involved, it is not always enough to say how we end up. It will sometimes be important to describe the exact process by which we are to reach our destination. (This seems likely to be true of most simple resistant two-way

exhibit 20 of chapter 19 -- Nebraska voting 1

A cautionary example
(all in 0.1%D)

A) The DATA -- from exhibit 3 of chapter 17

County	'20	'24	'28	'32
I2	373	368	332	838
H3	278	286	273	618
G4	299	336	251	612

B) ONE analysis with ZERO STRIPE MEDIANS*

					eff
I2	21	0	0	-27	53
H3	-8	-16	7	19	-13
G4	0	21	-28	0	0
eff	10	26	-10	321	289

*H3 median only zero by rounding, but matches panel C.

C) ANOTHER analysis with ZERO STRIPE MEDIANS*

I2	19	0	0	-29	55
H3	-9	-15	8	18	-12
G4	0	23	-26	0	0
eff	11	25	-11	322	288

*H3 median only zero by rounding, but matches panel C.

D) A THIRD analysis with ZERO STRIPE MEDIANS

I2	26	0	0	-22	48
H3	-5	-18	5	22	-16
G4	0	16	-33	0	0
eff	7	28	-8	318	292

E) A FOURTH analysis with ZERO STRIPE MEDIANS

I2	24	0	0	-24	50
H3	-6	-17	6	21	-15
G4	0	18	-31	0	0
eff	8	27	-9	319	291

analyses.) Now we know this, we need not be very concerned. All this means is:

When we need a unique answer, we are likely to have to go back to the data. It may not be enough to polish up any old analysis that is around.

The more important moral is not so clear until pointed out. Look again at exhibit 20. Do we care which of these analyses we have? Surely not for any exploratory purpose.

As the lawyers say: De minimis non curat lex. (The law pays no attention to matters that are too trivial.) As explorers who are we to do otherwise.

Clearly, then, we need not insist on uniqueness of exploratory analyses -- even by analysts with the same preconceptions and insights. It is enough if the possible analyses -- one of which we shall look at -- are reasonably close to one another.

This means more than merely "You can probably get along with any analysis with zero stripe medians". It means "You can probably get along with any analysis with 'small' stripe medians". The only problem is to define "small" in a reasonable way, preferably in a way that follows the scale of the irregularities shown by the data.

We already had this attitude in chapter 16, though we were not as explicit about it. Now it is time to be explicit about it and to use it to help us make further gains.